



# Estimation of a cumulative distribution function under interval censoring "case 1" via warped wavelets

Christophe Chesneau, Thomas Willer

## ► To cite this version:

Christophe Chesneau, Thomas Willer. Estimation of a cumulative distribution function under interval censoring "case 1" via warped wavelets. Communications in Statistics - Theory and Methods, 2015, 44 (17), 10.1080/03610926.2013.851231 . hal-00715260v4

**HAL Id: hal-00715260**

**<https://hal.science/hal-00715260v4>**

Submitted on 1 Dec 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of a cumulative distribution function under interval censoring “case 1” via warped wavelets

Christophe Chesneau<sup>1</sup> and Thomas Willer<sup>2</sup>

*Abstract:* The estimation of an unknown cumulative distribution function in the interval censoring “case 1” model from dependent sequences is considered. We construct a new adaptive estimator based on a warped wavelet basis and a hard thresholding rule. Under mild assumptions on the parameters of the model, considering the  $\mathbb{L}_2$  risk and the weighted Besov balls, we prove that the estimator attains a sharp rate of convergence. We also investigate its practical performances thanks to simulation experiments.

*Key words and phrases:* Adaptive estimation, Strongly mixing, Interval censoring, Warped wavelets, Hard thresholding.

*AMS 2000 Subject Classifications:* 62G05, 62G20.

## 1 Introduction

The mathematical context of the interval censoring “case 1” model can be described as follows: let  $(\delta_i, U_i)_{i \in \mathbb{Z}}$  be a strictly stationary process where, for any  $i \in \mathbb{Z}$ ,

$$\delta_i = \mathbf{1}_{\{X_i \leq U_i\}},$$

$\mathbf{1}_{\mathcal{A}}$  is the indicator function on any random event  $\mathcal{A}$ ,  $X_i$  and  $U_i$  are independent for any  $i$ , and  $(X_i)_{i \in \mathbb{Z}}$  is a strictly stationary process with common unknown cumulative distribution  $F$ . We assume that  $U_1$  admits a density, denoted by  $g$ , and we denote by  $G$  its cumulative distribution function. Our goal is to estimate  $F$  under mild assumptions on  $g$  from  $n$  observations  $(\delta_1, U_1), \dots, (\delta_n, U_n)$  of  $(\delta_i, U_i)_{i \in \mathbb{Z}}$ . This model has applications in Demography and Biology. See e.g. [14] and [19], and the references therein.

For recent statistical results, we refer to [25], [2] and [7]. In particular, considering the independent case, [2] have constructed adaptive penalized minimum contrast estimators built on trigonometric, polynomial or wavelet spaces. Using the  $\mathbb{L}_2$  risk over Besov balls, under some boundedness assumptions on  $g$ , [2, Corollary 3.1] proves that it attains the standard rate of convergence “ $n^{-2s/(2s+1)}$ ” where  $s$  characterizes the smoothness of  $F$ .

However, the independence assumption on  $(\delta_i, U_i)_{i \in \mathbb{Z}}$  is often stringent in applications. In this study, we investigate the adaptive estimation of  $F$  in a dependent setting (including the independent one). The so-called *strong mixing case* is considered. Examples and applications of this kind of dependence can be found in [4] and [16].

Assuming that  $g$  is known but with no boundedness assumptions on it, we develop a new adaptive estimator based on a warped wavelet basis and a hard thresholding rule. The features of this basis consist of a standard wavelet basis and of the definition of  $G$  related to the model. This enables us to give a significant stability to our thresholding algorithm. Such a technique has been already used with success in the framework of nonparametric regression with random design by [20]. Recent works on warped wavelet basis in nonparametric statistics can be found in [8], [9], [3], [24], [5] and [6].

---

<sup>1</sup>Université de Caen, LMNO, Campus II, Science 3, 14032 Caen France

<sup>2</sup>Aix Marseille Université, CNRS, Centrale Marseille, LATP, UMR 7353, 13453 Marseille France

Considering the  $\mathbb{L}_2$  risk over weighted Besov balls, we prove that our estimator attains the rate of convergence “ $(\ln n/n)^{2s/(2s+1)}$ ”, where  $s$  characterizes the smoothness of  $F$ . This rate of convergence corresponds to the one attained in the *i.i.d.* case (see [2]) up to an extra logarithmic term. Finally, we explore the numerical performances of the estimator.

The rest of the paper is organized as follows. Section 2 introduces notations and assumptions on the model. In Section 3, we describe warped wavelet bases on  $[0, 1]$  and weighted Besov balls. Our adaptive wavelet estimator is defined in Section 4. Theoretical and practical results are presented in Section 5. The proofs are postponed to Section 6.

## 2 Notations and assumptions

### 2.1 Assumptions on the dependence structure of the process

For any  $m \in \mathbb{Z}$ , we define the  $m$ -th strongly mixing coefficient of  $(X_i, U_i)_{i \in \mathbb{Z}}$  by

$$a_m = \sup_{(A,B) \in \mathcal{F}_{-\infty,0}^{(X,U)} \times \mathcal{F}_{m,\infty}^{(X,U)}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, \quad (2.1)$$

where  $\mathcal{F}_{-\infty,0}^{(X,U)}$  is the  $\sigma$ -algebra generated by the pairs of random variables  $\dots, (X_{-1}, U_{-1}), (X_0, U_0)$  and  $\mathcal{F}_{m,\infty}^{(X,U)}$  is the  $\sigma$ -algebra generated by the random variables  $(X_m, U_m), (X_{m+1}, U_{m+1}), \dots$

We consider the exponentially strongly mixing case: there exist two constants  $\gamma > 0$  and  $c > 0$  such that, for any integer  $m \geq 1$ ,

$$a_m \leq \gamma \exp(-cm). \quad (2.2)$$

This assumption is not very restrictive; some examples of processes satisfying such conditions can be found in e.g. [26], [16], [23] and [4].

### 2.2 Assumptions on the densities

For the sake of simplicity, we suppose that all the considered random variables take their values in  $[0, 1]$ .

In the main part of the study, we assume that  $g$  is known. The unknown case will only be explored in the simulation study in subsection 6.2.

We suppose that, for any interval  $[a, b] \subseteq [0, 1]$ , there exists a constant  $C > 0$  such that

$$\left( \frac{1}{b-a} \int_a^b g(x)^2 dx \right)^{1/2} \leq C \frac{1}{b-a} \int_a^b g(x) dx. \quad (2.3)$$

This “reverse Hölder inequality” is related to the Muckenhoupt weights theory. It includes a wide variety of densities, non-necessarily bounded from above and/or below. For instance,  $g(x) = (u+1)x^u$ ,  $x \in [0, 1]$  and  $u \in (0, 1)$  satisfies (2.3). Further details can be found in [20, subsection 4.1].

For any  $m \in \{1, \dots, n\}$ , let  $\mathbf{f}_{(X_0, U_0, X_m, U_m)}$  be the density of  $(X_0, U_0, X_m, U_m)$ ,  $\mathbf{f}_{(X_0, U_0)}$  the density of  $(X_0, U_0)$  and, for any  $(y, x, y_*, x_*) \in [0, 1]^4$ ,

$$h_m(y, x, y_*, x_*) = \mathbf{f}_{(X_0, U_0, X_m, U_m)}(y, x, y_*, x_*) - \mathbf{f}_{(X_0, U_0)}(y, x) \mathbf{f}_{(X_0, U_0)}(y_*, x_*). \quad (2.4)$$

We suppose that there exists a constant  $C > 0$  such that

$$\sup_{m \in \{1, \dots, n\}} \sup_{(x, x_*) \in [0, 1]^2} \frac{1}{g(x)g(x_*)} \int_0^x \int_0^{x_*} |h_m(y, x, y_*, x_*)| dy dy_* \leq C. \quad (2.5)$$

Note that, in the independent case, we have  $h_m(y, x, y_*, x_*) = 0$  and (2.5) is satisfied. Moreover, functions  $g$  satisfying (2.3) and (2.5) are not necessarily bounded from below and above. Hence our conditions are less restrictive than [2, Assumption A1].

### 3 Warped wavelets and weighted Besov balls

Let  $N$  be a positive integer. We consider an orthonormal wavelet basis generated by dilations and translations of a "father" Daubechies-type wavelet  $\phi$  and a "mother" Daubechies-type wavelet  $\psi$  of the family  $db2N$  (see [11]). In particular, mention that  $\phi$  and  $\psi$  have compact supports.

We set

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Suppose that (2.3) holds and recall that

$$G(x) = \mathbb{P}(U_1 \leq x) = \int_0^x g(u)du, \quad x \in \mathbb{R}.$$

Then, with an appropriate treatment at the boundaries, there exists an integer  $\tau$  satisfying  $2^\tau \geq 2N$  such that, for any integer  $j_* \geq \tau$ , any  $h \in \mathbb{L}_2([0, 1]) = \left\{ h : [0, 1] \rightarrow \mathbb{R}; \int_0^1 h^2(x)dx < \infty \right\}$  can be expanded into a warped wavelet series as

$$h(x) = \sum_{k=0}^{2^{j_*}-1} \alpha_{j_*,k} \phi_{j_*,k}(G(x)) + \sum_{j=j_*}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(G(x)), \quad x \in [0, 1],$$

where

$$\alpha_{j,k} = \int_0^1 h(G^{-1}(x)) \phi_{j,k}(x) dx, \quad \beta_{j,k} = \int_0^1 h(G^{-1}(x)) \psi_{j,k}(x) dx. \quad (3.1)$$

See [20, subsection 3.3].

Let  $M > 0$  and  $s > 0$ . We say that a function  $h$  in  $\mathbb{L}_2([0, 1])$  belongs to the weighted Besov ball  $B_{s,\infty}^w(M)$  if there exists a constant  $M > 0$  such that the associated wavelet coefficients (3.1) satisfy, for any integer  $j \geq \tau$ ,

$$\sum_{k=0}^{2^j-1} \beta_{j,k}^2 w_{j,k} \leq M 2^{-j(2s+1)},$$

where

$$w_{j,k} = \int_{k/2^j}^{(k+1)/2^j} \frac{1}{g(G^{-1}(x))} dx. \quad (3.2)$$

In this expression,  $s$  is a smoothness parameter. Details concerning the warped wavelets and the analytic definition of weighted Besov balls can be found in [20, Section 7]. For the standard wavelet basis on  $[0, 1]$ , see e.g. [22] and [10].

### 4 Estimators

For any integer  $j \geq \tau$  and any  $k \in \{0, \dots, 2^j - 1\}$ , we estimate the unknown warped wavelet coefficients of  $F$  i.e.  $\alpha_{j,k} = \int_0^1 F(G^{-1}(x)) \phi_{j,k}(x) dx$  and  $\beta_{j,k} = \int_0^1 F(G^{-1}(x)) \psi_{j,k}(x) dx$  by respectively

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^n \delta_i \phi_{j,k}(G(U_i)), \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^n \delta_i \psi_{j,k}(G(U_i)). \quad (4.1)$$

Some of their statistical properties are investigated in Propositions 6.1, 6.2 and 6.3 below.

We estimate  $F$  by the following hard thresholding estimator  $\hat{F}$ :

$$\hat{F}(x) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k} \phi_{j_0,k}(G(x)) + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa \rho_n\}} \psi_{j,k}(G(x)), \quad (4.2)$$

where  $x \in [0, 1]$ ,  $j_0$  is an integer such that

$$\frac{1}{2} \ln n < 2^{j_0} \leq \ln n,$$

$\hat{\alpha}_{j_0,k}$  and  $\hat{\beta}_{j,k}$  are defined by (4.1),  $j_1$  is the integer satisfying

$$\frac{1}{2} \frac{n}{(\ln n)^3} < 2^{j_1} \leq \frac{n}{(\ln n)^3}, \quad (4.3)$$

$\kappa$  is a large enough constant (the one in Proposition 6.3 below) and  $\rho_n$  denotes the “universal threshold”, i.e.

$$\rho_n = \sqrt{\frac{\ln n}{n}}. \quad (4.4)$$

Naturally, for  $x < 0$ , we put  $\hat{F}(x) = 0$  and, for  $x > 1$ ,  $\hat{F}(x) = 1$ .

Note that  $\hat{F}$  is adaptive: its construction does not depend on the smoothness of  $F$ .

The general idea in the construction of  $\hat{F}$  is to apply a term-by-term selection on the unknown wavelet coefficients of  $F$ : only the most significant are kept. The reason is that only these few coefficients contain the main characteristics of  $F$ . For the construction of hard thresholding wavelet estimators in the standard nonparametric models, see e.g. [15], [13] and [18].

When  $g$  is unknown, an intuitive adaptive estimator of  $F$  is (4.2) but, instead of  $G$ , we consider its empirical version:

$$\hat{G}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i \leq x\}}.$$

This plug-in method yields the hard thresholding estimator

$$\hat{F}^*(x) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}^* \phi_{j_0,k}(\hat{G}_n(x)) + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k}^* \mathbf{1}_{\{|\hat{\beta}_{j,k}^*| \geq \kappa \rho_n\}} \psi_{j,k}(\hat{G}_n(x)), \quad (4.5)$$

$x \in [0, 1]$ , where

$$\hat{\alpha}_{j,k}^* = \frac{1}{n} \sum_{i=1}^n \delta_i \phi_{j,k}(\hat{G}_n(U_i)), \quad \hat{\beta}_{j,k}^* = \frac{1}{n} \sum_{i=1}^n \delta_i \psi_{j,k}(\hat{G}_n(U_i)).$$

We do not investigate the theoretical performances of this estimator, but we use it throughout the simulation and real data study of Section 5.2.

## 5 Performances of $\hat{F}$

### 5.1 Theoretical results

**Theorem 5.1** *Suppose that the assumptions of Section 2 hold. Let  $\hat{F}$  be (4.2). Suppose that  $F \in B_{s,\infty}^w(M)$  with  $s > 0$ . Then there exists a constant  $C > 0$  such that*

$$\mathbb{E} \left( \int_0^1 \left( \hat{F}(x) - F(x) \right)^2 dx \right) \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 5.1 uses a suitable decomposition of the  $\mathbb{L}_2$  risk including some geometrical properties of the warped wavelet basis and the statistical properties of the wavelet coefficients estimators presented in Propositions 6.1, 6.2 and 6.3.

Theorem 5.1 shows that, under mild assumptions on the dependence of the observations and on  $g$ ,  $\hat{F}$  attains a rate of convergence close to the one for the *i.i.d.* case i.e.  $n^{-2s/(2s+1)}$ . The difference is the “negligible” logarithmic term  $(\ln n)^{2s/(2s+1)}$ .

Let us recall that, if we restrict our study to the independent case, contrary to [2, Assumption A1], Theorem 5.1 holds without boundedness assumptions on  $g$ .

## 5.2 Practical results

This section is devoted to the numerical performances of  $\hat{F}$ . We restrict ourselves to the *i.i.d.* case: no exponentially strongly mixing hypothesis is made here.

### 5.2.1 Preliminary remarks on the estimators

In practice, the construction and the performances of the two estimators in the case of known or unknown design density are very close. Thus for the sake of brevity, we present only the results for the unknown design density procedure.

We compare the performances of the warped wavelet estimator to those of the piecewise polynomial regression estimator developed in [2]. Another procedure, named quotient estimator, is also proposed in that paper. However it is generally outperformed by the regression estimator, so we focus only on the warped wavelet and the piecewise polynomial estimators. *We name them respectively WWE and PPE in the sequel.* We give examples of their behaviours when applied to various simulated data sets and to one real data set.

In both methods, one needs to fit several preliminary parameters. For the WWE, the calibration of the threshold  $\kappa\rho_n$  and of the cutoff level  $j_1$  are important practical issues. The values given by the theory are not useful in practice. Indeed the thresholds are defined up to some intricate “large enough constant  $\kappa$ ”. Moreover the cutoff level  $j_1$  defined in (4.3) is too small in practice, as it is even lower than the minimal coarsest level value. Thus, for  $j_1$ , we use  $j_1 = \log_2(n) - 1$ , i.e. the maximal possible level, hoping the high resolution “noise” is filtered thanks to thresholding.

For the thresholds, we can use an analogy with the non-parametric regression model in random design. Indeed the censoring model amounts to observing at each design point  $U_i$  some Bernoulli variable  $\delta_i$  with (conditionally to  $U_i$ ) expectation  $F(U_i)$  and standard deviation  $\sqrt{F(U_i)(1 - F(U_i))}$ . In the usual regression model one observes at each design point  $U_i$  some Gaussian variable  $\delta_i$  with (conditionally to  $U_i$ ) expectation  $F(U_i)$  and some given standard error  $\sigma$ . Thus we calibrate  $\kappa$  as in regression (see [9]) as the following random value:

$$\kappa_r = \sqrt{\frac{1}{n-2} \sum_{i=2}^n (\delta_{(i)} - \delta_{(i-1)})^2},$$

where each  $\delta_{(i)}$  refers to the value  $\delta_k$  such that  $U_k$  is the  $i$ -th higher coordinate of the vector  $(U_j)_{1 \leq j \leq n}$ . Nevertheless, the standard error is highly heteroscedastic in the censoring problem: it varies from 0 for observations near the edges of the observation interval, to 0.5 near the median of  $F$ . Thus we also use a more conservative deterministic threshold corresponding to

$$\kappa_d = 0.5 \times \sqrt{2} = \frac{1}{\sqrt{2}}.$$

Thus, we make sure that the high noise of the median observations is well taken into account in the thresholds. *In the sequel, we name the random threshold estimator WWER and the deterministic threshold estimator WWEd.*

Lastly, we use the 'Daubechies 4' wavelet basis for the simulated data study, and the 'Daubechies 3' wavelet basis for the real data study (as the number of observations is a bit small for Daubechies 4 basis).

The PPE also requires to fix some preliminary parameters. First some maximal degree of the polynomials and some maximal numbers of intervals must be put as arguments. We choose the degree as equal to 6 and maximum number of intervals as equal to 64 for the simulated data, and to 32 for the real data. Secondly some noise variance level must be put as an argument as well. We make two choices corresponding to the random and deterministic values used for WWEr and WWEd. *Likewise, in the sequel, we name the corresponding estimators PPEr and PPEd.*

Hence for each data set we compare the performances of four estimators: WWEr, WWEd, PPEr and PPEd.

Concerning the programs, the WWE is very simple to implement in practice. The program consists mainly in simple wavelet decompositions or recompositions, computations of thresholds, and shrinkage of some small wavelets coefficients. The PPE was implemented thanks to the FY3P.m Matlab file and several other sub-programs (all available on Yves Rozenholc's web page). The implementation is much more complex, as can be seen from the programs. What is more, the processing times necessary to compute the estimators are very different. For example one needs a CPU time of 0.078 to compute the eight Warped Wavelets Estimators represented in Figure 5.2.2, whereas one needs a CPU time of 670.49 to compute the eight Piecewise Polynomial Estimators built on the same data and represented in the same figure.

## 5.2.2 Study with simulated data

### Simulation of the data

We consider four target functions  $F$  which correspond to the following formula with values  $q = 0.2, 1, 5, 50$ :

$$F_q(t) = \begin{cases} 2^{q-1}t^q, & t \in [0, 0.5], \\ 1 - 2^{q-1}(1-t)^q, & t \in [0.5, 1]. \end{cases} \quad (5.1)$$

These four functions are plotted in Figure 1. Furthermore we consider five different densities of the  $U_i$ s. We name them Bump1, Bump2, Pit1, Pit2 and Uniform. They correspond to the following formulas (up to normalization constants), and are plotted in Figure 2.

- Bump1  $g(x) = \exp(-(100 * (x - 0.5)^2)) + 1$ ,
- Pit1  $g(x) = -\exp(-(100 * (x - 0.5)^2)) + 1.05$ ,
- Bump2  $g(x) = \exp(-(100 * x^2)) + \exp(-(100 * (x - 1)^2)) + 1$ ,
- Pit2  $g(x) = -\exp(-(100 * x^2)) - \exp(-(100 * (x - 1)^2)) + 1.05$ ,
- Uniform  $g(x) = 1$ .

The non uniform densities correspond to a surplus or a lack of design points near the edges or near the middle of the interval. The WWE should normally capture detail coefficients in zones of dense design, and fail to do so in low design density zones. Nevertheless, this design effect can be counterbalanced by the variations of the level of noise in the observation, as mentioned previously. Thus it is difficult to predict the performances of the estimator, especially in the middle of the interval. As an example, the wavelet detail coefficients of the targets are plotted in Figure 3. One can remark that  $F_q$  has high detail coefficients near the edges of the interval  $[0, 1]$  when  $q$  is small, and at the middle of the interval  $[0, 1]$  when  $q$  is large. One can expect estimation problems for example for the  $F_5$  or  $F_{50}$  functions in the case of the 'Pit1' design density.

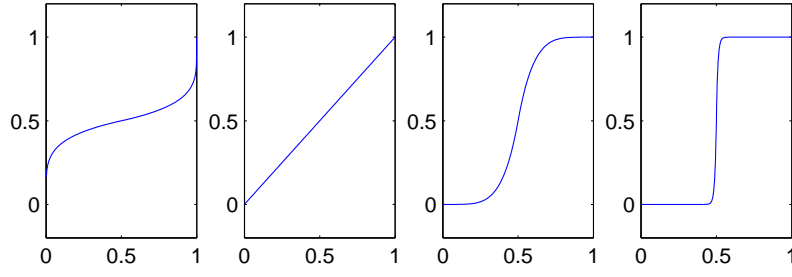


Figure 1: Four target cumulative distribution functions:  $F_{0.2}$ ,  $F_1$ ,  $F_5$ ,  $F_{50}$  (from left to right)

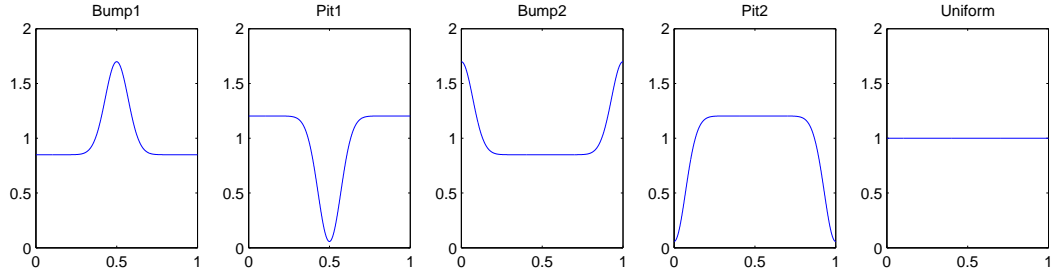


Figure 2: Densities of the  $(U_i)$ s: Bump1, Pit1, Bump2, Pit2 and Uniform (from left to right)

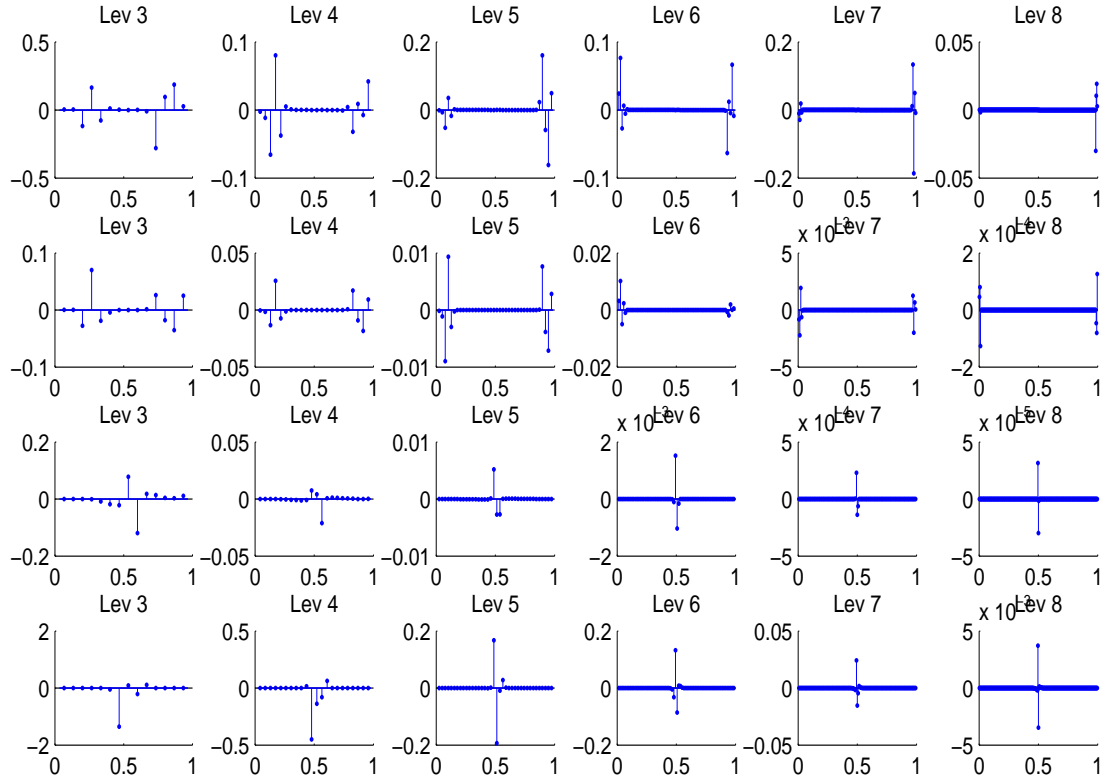


Figure 3: Detail wavelet coefficients of each target function: the first row relates to  $F_{0.2}$ , the second to  $F_1$ , the third to  $F_5$  and the fourth to  $F_{50}$ .



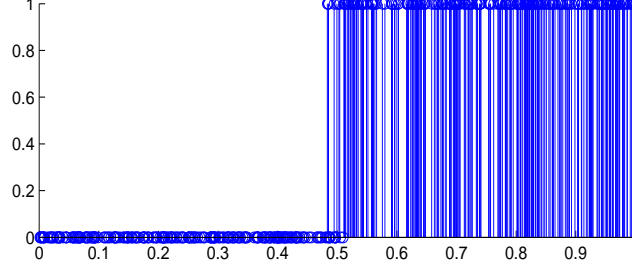


Figure 4: Data with  $(U_i)$ s (uniformly distributed) on axis 1, and their corresponding  $(\delta_i)$ s on axis 2, for the  $F_{50}$  function.

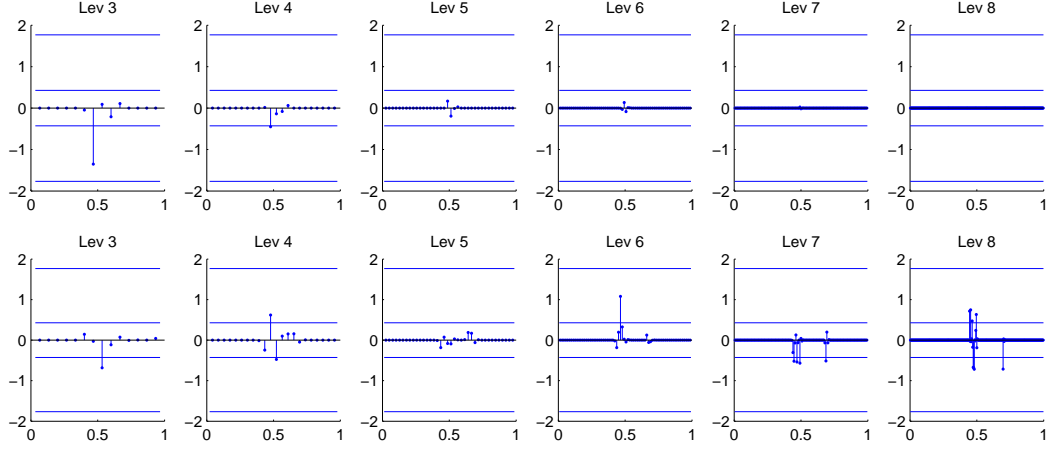


Figure 5: Random thresholds (lower horizontal bars), deterministic thresholds (higher horizontal bars), wavelet detail coefficients of  $F_{50}$  (first row of graphics), wavelet detail coefficients of the  $(\delta_i)$ s (second row of graphics)

### An example of data and wavelet estimator

An example of the data on which the estimators are based is given in Figures 4 and 5 (for the cdf  $F_{50}$  and for a uniform distribution of the  $U_i$ s). Figure 4 represents the couples  $(U_i, \delta_i)$ . The second row of Figure 5 gives the detail wavelet coefficients of the sequence  $(\delta_{\sigma_U(i)})_{i \in \{1, \dots, n\}}$ , where  $\sigma_U$  is the following permutation:  $\sigma_U(i)$  is the index  $j$  such that  $U_j$  is the  $i^{st}$  largest element of the sequence  $U$ . These coefficients are approximations of the ones given in equation (4.1), as their construction consists in replacing  $G$  by the empirical pdf  $\hat{G}_n$  in their expressions. They should be estimations of the coefficients of the target function, which are plotted in the first row of Figure 5. Then the estimation consists in trying to select the coefficients corresponding to the signal  $F$ , and to put to zero those corresponding to the "noise".

As one can see from Figure 5, the thresholding is a hard task as the coefficients of the data are significantly different from the target coefficients. This may stem from the high noise of the observations in the middle of the interval, as discussed previously. The deterministic threshold is conservative as all the detail coefficients are thresholded. The random threshold is too low and leaves high resolution noise unfiltered. The corresponding estimators can be seen in Figure 6, and WWER indeed contains artefacts in the middle of the interval.

### Some examples of results

We give examples of realizations of the 4 estimators for all the 20 simulated data sets. Figures 6, 7, 8, 9, 10 give examples for respectively the Uniform, Bump1, Pit1, Bump2, and Pit2 design densities. Each figure represents the target function (dashed line) and the estimator (solid line). We can draw the following conclusions.

- All the estimators behave rather properly at the edges of the interval, while there are sometimes problems in the middle of the interval. This is coherent with the variations of the noise variance, as discussed previously. As one could expect, the random threshold estimators have poor performances in the middle of the interval for the Uniform and Pit1 designs.
- Generally, the deterministic threshold estimators are better than the random threshold estimators. Probably the latter underestimate the noise level in the middle of the interval, and thus their performances are poor, especially when the number of observations in this zone is small.
- For the  $F_{0.2}$  or  $F_1$  targets, all the estimators are generally rather close. One can remark however that PPE is a bit better than WWE. One can check that they consist of rough approximation methods: for WWE all the detail coefficients are cut, and for PPE low degree polynomials and few intervals are selected. For WWE, one may ask oneself if the results could be improved by using other calibrations of  $\kappa$  and  $j_1$ . For this purpose, one can compute some "oracle" estimator in the sense that it uses the threshold and the cutoff values which minimize the mean square error among all estimators of the type of Section 4, except with  $\kappa$  and  $j_1$  left free. Of course this estimator is completely inaccessible in practice. Monte Carlo approximations of the mean square error show that the oracle strategy consists in putting all the detail coefficients to zero. Thus these two functions probably lead to globally noisy observations, and it must be hopeless to try to recover details of the targets in these cases.
- For the  $F_5$  target, WWE generally outperforms PPE. However, for the  $F_{50}$  target, PPE generally outperforms WWE.

### 5.2.3 Study with real data

We use a data set concerning a tumorigenicity experiment conducted by the National Toxicology Program (NTP) which is described and summarized in Dunson and Dinse (2002). There are four sets of data. We use only the one corresponding to the 'control' group and the presence of adrenal tumor. The subjects were fifty rats or mice. The  $X_i$ s are unobserved dates of occurrence of an adrenal tumor and the  $U_i$  are dates of death. Every month during 25 months, all the dead mice or rats were examined to check if they had a tumor or not, which corresponds to respectively  $\delta_i = 1$  and  $\delta_i = 0$ . The experience stopped after the 25th month (we consider this date as the 26th month) when all the remaining were killed and examined.

Several remarks need to be made on this data set. First at month 26 the data show that 5 out of the 13 remaining mice had no tumor, thus the upper bound of the support of  $F$  was not reached. This is why in the sequel, our estimators stop at about 0.7 and not 1. Secondly, the design dates cannot be considered as realizations of continuous random variables. In particular, since the observations are gathered by months, several deaths occurred at exactly the same time. Thus we considered that all the deaths at month  $m$  were in fact distributed on a regular grid between months  $m - 1$  and  $m$ .

The four estimators are plotted in Figure 11. The two thresholding techniques yield to the same estimator. The PPE yields a rough one degree polynomial estimation, whereas the WWE yields a more detailed curve.

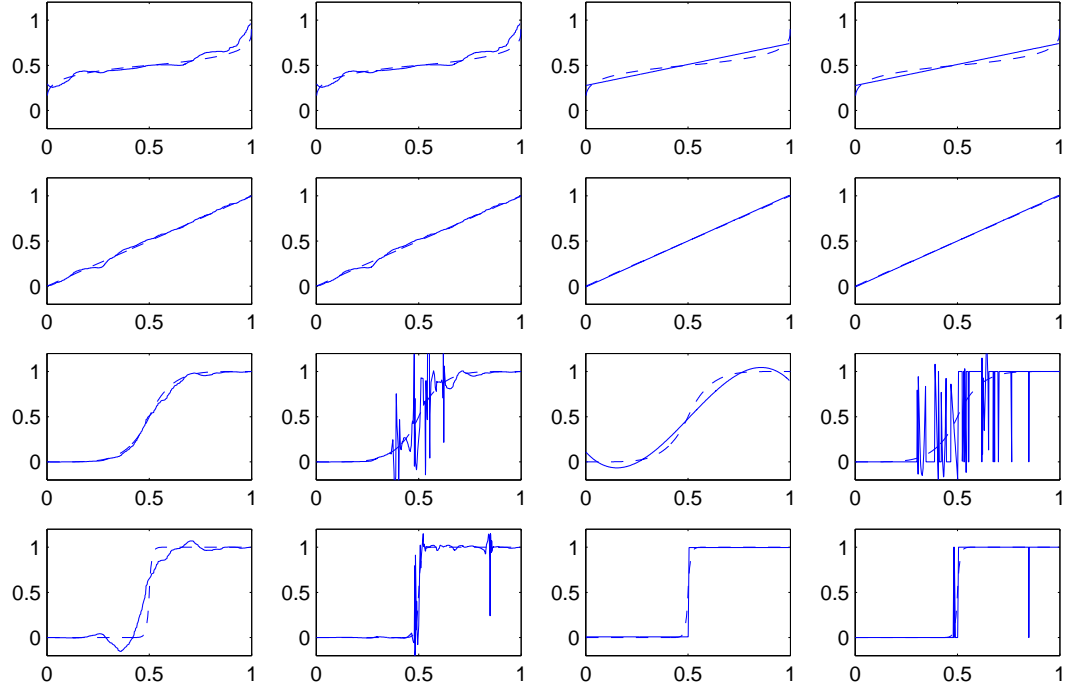


Figure 6: Uniform design: estimators WWEd, WWER, PPed, PPER (left to right) for targets  $F_{0.2}$ ,  $F_1$ ,  $F_5$ ,  $F_{50}$  (from top to bottom)

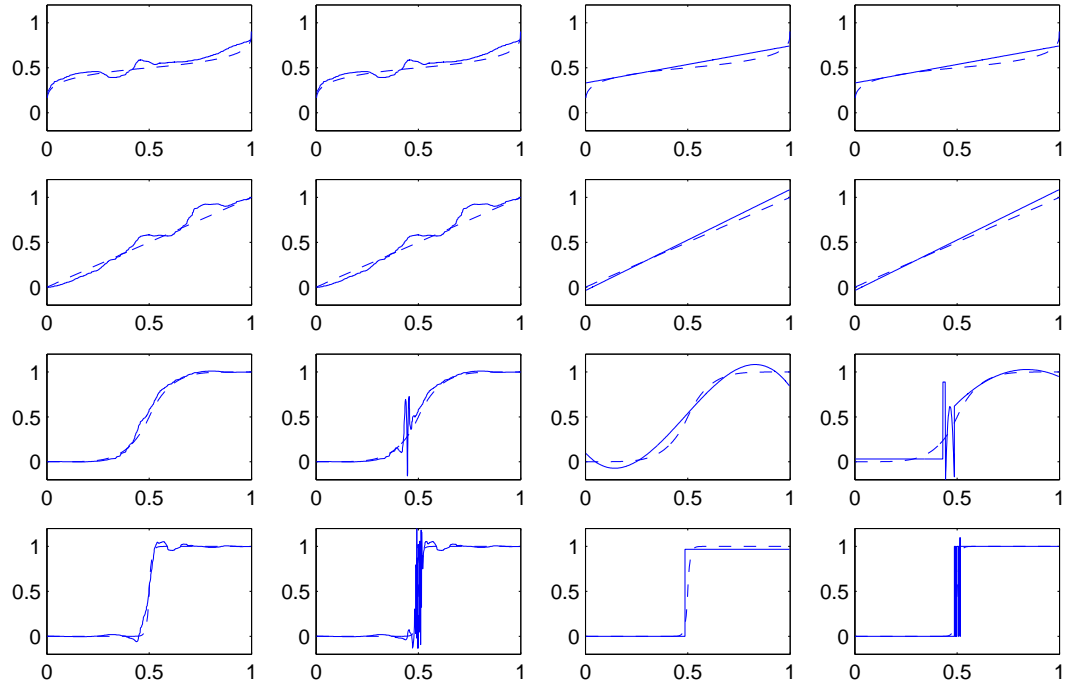


Figure 7: Bump1 design: estimators WWEd, WWER, PPed, PPER (left to right) for targets  $F_{0.2}$ ,  $F_1$ ,  $F_5$ ,  $F_{50}$  (from top to bottom)

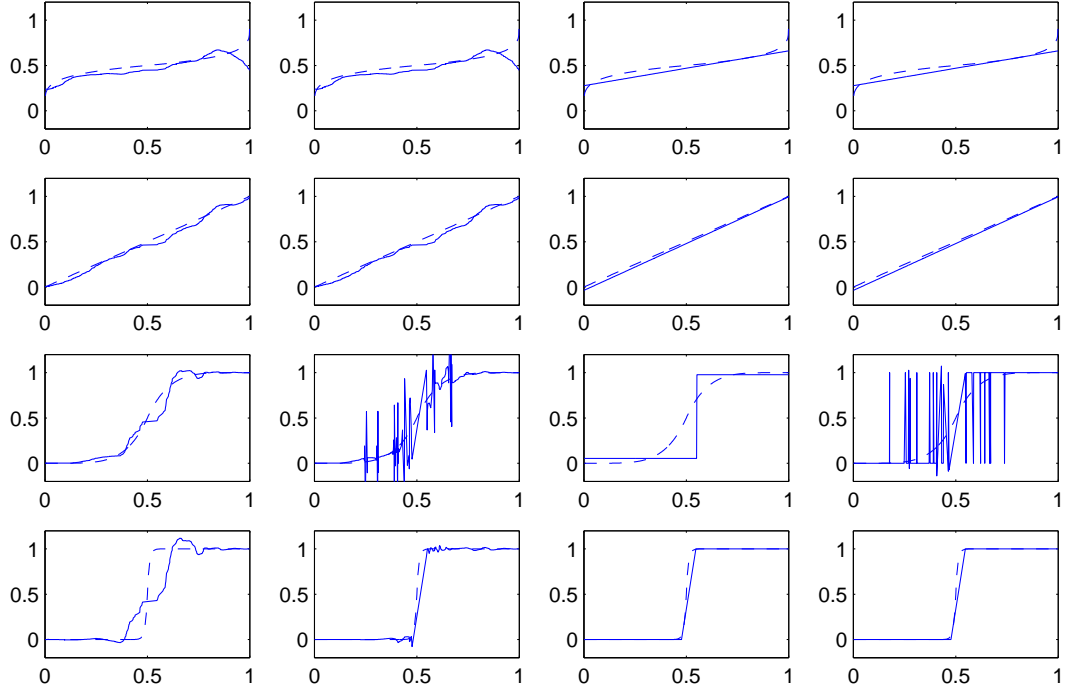


Figure 8: Pit1 design: estimators WWEd, WWEr, PPed, PPer (left to right) for targets  $F_{0.2}$ ,  $F_1$ ,  $F_5$ ,  $F_{50}$  (from top to bottom)

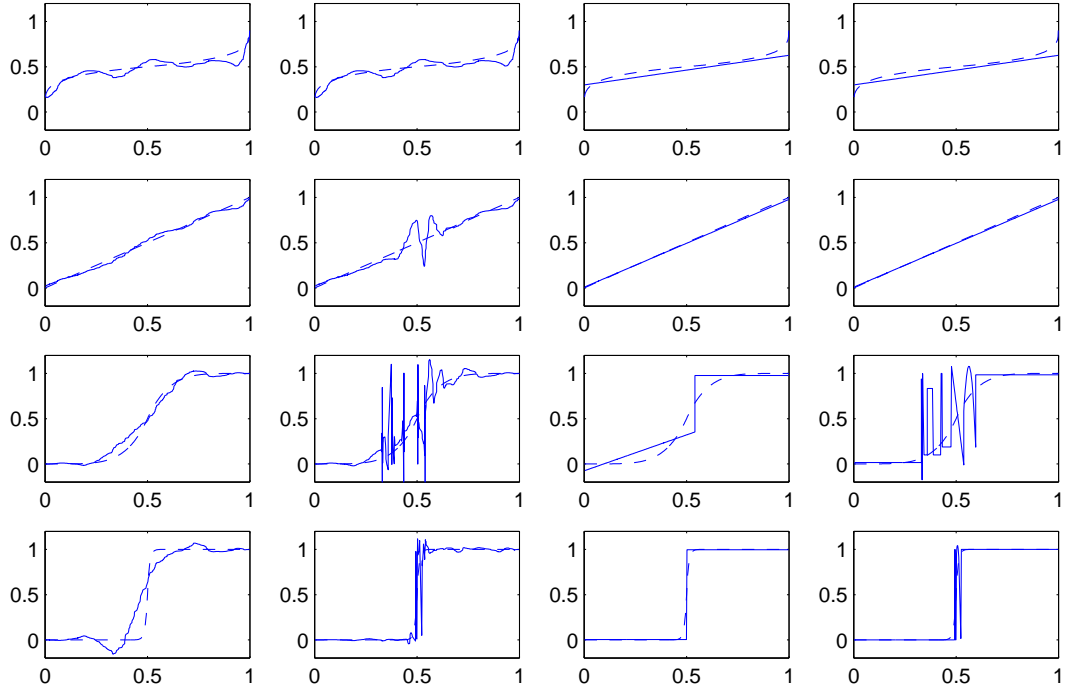


Figure 9: Bump2 design: estimators WWEd, WWEr, PPed, PPer (left to right) for targets  $F_{0.2}$ ,  $F_1$ ,  $F_5$ ,  $F_{50}$  (from top to bottom)

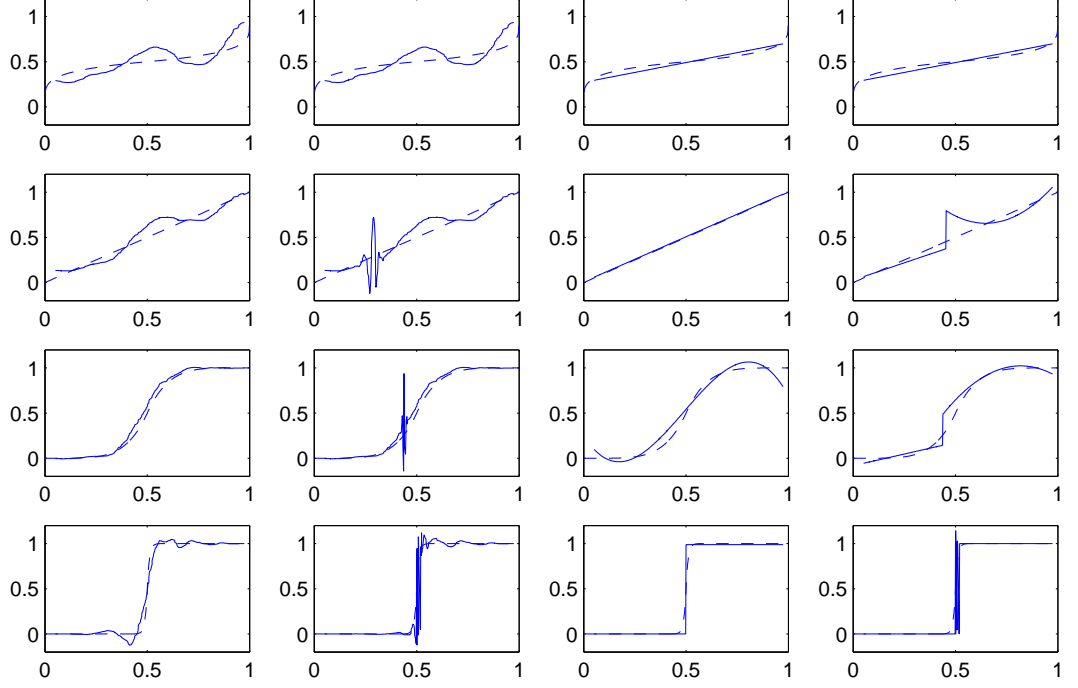


Figure 10: Pit2 design: estimators WWEd, WWER, PPed, PPER (left to right) for targets  $F_{0.2}$ ,  $F_1$ ,  $F_5$ ,  $F_{50}$  (from top to bottom)

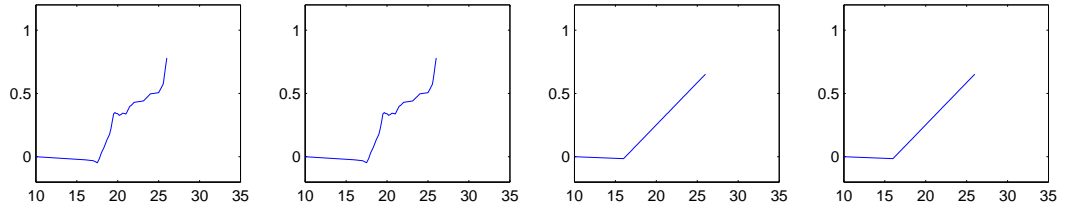


Figure 11: Estimators WWEd, WWER, PPed, PPER (left to right) for the NTP data set

### 5.3 Conclusion and discussion

In this paper we develop a new adaptive estimator for the cumulative distribution function under interval censoring “case 1” for possible dependent data. It is constructed from warped wavelet basis and a hard thresholding rule. Theoretical results show the good performance of our estimator under mild assumptions on the model (including vanishing density  $g$ ). The practical performances were investigated thanks to a comparison with the estimator developed in [2]. The wavelet estimator displays generally satisfying performances, provided a conservative calibration of the threshold is made. Moreover it is simpler and faster to compute than the estimator of [2].

Possible perspectives of this work are to

- determine the rate of convergence of  $\hat{F}^*$  (4.5) under the  $\mathbb{L}_2$  risk over Besov balls,
- relax assumptions (2.2) and/or (2.5),
- improve the obtained rate of convergence by considering more sophisticated thresholding rules as those developed in [1].

## 6 Proofs

In this section, we suppose that the assumptions of Section 2 hold. Moreover,  $C$  denotes any constant that does not depend on  $j$ ,  $k$  and  $n$ . Its value may change from one term to another and may depend on  $\phi$  or  $\psi$ .

### 6.1 Auxiliary results on (4.1)

**Proposition 6.1** *Suppose that the assumptions of Section 2 hold. For any integer  $j \geq j_0$  such that  $2^j \leq n$  and any  $k \in \{0, \dots, 2^j - 1\}$ , let  $\alpha_{j,k} = \int_0^1 F(G^{-1}(x))\phi_{j,k}(x)dx$  and  $\hat{\alpha}_{j,k}$  be defined as in (4.1). Then  $\hat{\alpha}_{j,k}$  is an unbiased estimator for  $\alpha_{j,k}$  and there exists a constant  $C > 0$  such that*

$$\text{Var}(\hat{\alpha}_{j,k}) \leq C \frac{1}{n},$$

Let us mention that Proposition 6.1 can be proved with  $\hat{\beta}_{j,k}$  (4.1) instead of  $\alpha_{j,k}$  and  $\beta_{j,k} = \int_0^1 F(G^{-1}(x))\psi_{j,k}(x)dx$  instead of  $\alpha_{j,k}$ .

**Proof of Proposition 6.1.** We have

$$\begin{aligned} \mathbb{E}(\hat{\alpha}_{j,k}) &= \mathbb{E}(\delta_1 \phi_{j,k}(G(U_1))) = \mathbb{E}(\mathbf{1}_{\{X_i \leq U_i\}} \phi_{j,k}(G(U_1))) \\ &= \int_0^1 \int_0^u f(x) \phi_{j,k}(G(u)) g(u) dx du = \int_0^1 F(u) \phi_{j,k}(G(u)) g(u) du \\ &= \int_0^1 F(G^{-1}(u)) \phi_{j,k}(u) du = \alpha_{j,k}. \end{aligned}$$

An elementary covariance inequality yields

$$\text{Var}(\hat{\alpha}_{j,k}) = \frac{1}{n^2} \sum_{v=1}^n \sum_{\ell=1}^n \text{Cov}(\delta_v \phi_{j,k}(G(U_v)), \delta_\ell \phi_{j,k}(G(U_\ell))) \leq \mathbf{S} + \mathbf{T}, \quad (6.1)$$

where

$$\mathbf{S} = \frac{1}{n} \text{Var}(\delta_1 \phi_{j,k}(G(U_1)))$$

and

$$\mathbf{T} = \frac{2}{n^2} \left| \sum_{v=2}^n \sum_{\ell=1}^{v-1} \text{Cov} (\delta_v \phi_{j,k}(G(U_v)), \delta_\ell \phi_{j,k}(G(U_\ell))) \right|.$$

Since  $\delta_1 \leq 1$ , we have

$$\begin{aligned} \mathbf{S} &\leq \frac{1}{n} \mathbb{E} \left( (\delta_1 \phi_{j,k}(G(U_1)))^2 \right) \leq \frac{1}{n} \mathbb{E} \left( (\phi_{j,k}(G(U_1)))^2 \right) \\ &= \frac{1}{n} \int_0^1 (\phi_{j,k}(G(x)))^2 g(x) dx = \frac{1}{n} \int_0^1 (\phi_{j,k}(x))^2 dx = \frac{1}{n}. \end{aligned} \quad (6.2)$$

It follows from the stationarity of  $(\delta_i, U_i)_{i \in \mathbb{Z}}$  and  $2^j \leq n$  that

$$\begin{aligned} \mathbf{T} &= \frac{2}{n^2} \left| \sum_{m=1}^n (n-m) \text{Cov} (\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m))) \right| \\ &\leq \frac{2}{n} \sum_{m=1}^n |\text{Cov} (\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m)))| = \mathbf{T}_1 + \mathbf{T}_2, \end{aligned} \quad (6.3)$$

where

$$\mathbf{T}_1 = \frac{2}{n} \sum_{m=1}^{2^j-1} |\text{Cov} (\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m)))|$$

and

$$\mathbf{T}_2 = \frac{2}{n} \sum_{m=2^j}^n |\text{Cov} (\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m)))|.$$

*Upper bound for  $\mathbf{T}_1$ .* Using (2.4), (2.5) and the change of variables  $y = 2^j x - k$ , we obtain

$$\begin{aligned} &|\text{Cov} (\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m)))| \\ &= \left| \int_0^1 \int_0^1 \int_0^1 \int_0^1 h_m(y, x, y_*, x_*) (\mathbf{1}_{\{y \leq x\}} \phi_{j,k}(G(x)) \mathbf{1}_{\{y_* \leq x_*\}} \phi_{j,k}(G(x_*))) dy dx dy_* dx_* \right| \\ &\leq \int_0^1 \int_0^{x_*} \int_0^1 \int_0^x |h_m(y, x, y_*, x_*)| |\phi_{j,k}(G(x)) \phi_{j,k}(G(x_*))| dy dx dy_* dx_* \\ &\leq C \left( \int_0^1 |\phi_{j,k}(G(x))| g(x) dx \right)^2 \leq C \left( \int_0^1 |\phi_{j,k}(x)| dx \right)^2 \leq C 2^{-j}. \end{aligned}$$

Therefore

$$\mathbf{T}_1 \leq C \frac{1}{n} 2^{-j} 2^j = C \frac{1}{n}. \quad (6.4)$$

*Upper bound for  $\mathbf{T}_2$ .* Applying the Davydov inequality for strongly mixing processes (see [12]), for any  $q \in (0, 1)$ , we have

$$\begin{aligned} &|\text{Cov} (\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m)))| \\ &\leq C a_m^q \left( \mathbb{E} \left( |\delta_0 \phi_{j,k}(G(U_0))|^{2/(1-q)} \right) \right)^{1-q} \leq C a_m^q \left( \mathbb{E} \left( |\phi_{j,k}(G(U_0))|^{2/(1-q)} \right) \right)^{1-q} \\ &\leq C a_m^q \left( \sup_{x \in [0,1]} |\phi_{j,k}(G(x))| \right)^{2q} \left( \mathbb{E} \left( (\phi_{j,k}(G(U_0)))^2 \right) \right)^{1-q}. \end{aligned}$$

We have  $\sup_{x \in [0,1]} |\phi_{j,k}(x)| \leq C 2^{j/2}$  and, by (6.2),

$$\mathbb{E} \left( (\phi_{j,k}(G(U_0)))^2 \right) \leq 1.$$

Therefore

$$|\text{Cov}(\delta_0 \phi_{j,k}(G(U_0)), \delta_m \phi_{j,k}(G(U_m)))| \leq C 2^{qj} a_m^q.$$

Hence

$$\mathbf{T}_2 \leq C \frac{1}{n} 2^{qj} \sum_{m=2^j}^n a_m^q \leq C \frac{1}{n} \sum_{m=2^j}^n m^q a_m^q \leq C \frac{1}{n}. \quad (6.5)$$

It follows from (6.3), (6.4) and (6.5) that

$$\mathbf{T} \leq C \frac{1}{n}. \quad (6.6)$$

Combining (6.1), (6.2) and (6.6), we obtain

$$\text{Var}(\hat{\alpha}_{j,k}) \leq C \frac{1}{n}.$$

The proof of Proposition 6.1 is complete. □

**Proposition 6.2** *Suppose that the assumptions of Section 2 hold. For any integer  $j \geq j_0$  such that  $2^j \leq n$  and any  $k \in \{0, \dots, 2^j - 1\}$ , let  $\beta_{j,k} = \int_0^1 F(G^{-1}(x)) \psi_{j,k}(x) dx$  and  $\hat{\beta}_{j,k}$  be defined as in (4.1). Then there exists a constant  $C > 0$  such that*

$$\mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^4 \right) \leq C 2^j \frac{1}{n}.$$

**Proof of Proposition 6.2.** Observe that

$$\begin{aligned} |\hat{\beta}_{j,k} - \beta_{j,k}| &\leq |\hat{\beta}_{j,k}| + |\beta_{j,k}| \leq \sup_{(x,y) \in [0,1]^2} |y \psi_{j,k}(G(x))| + C \\ &\leq \sup_{x \in [0,1]} |\psi_{j,k}(x)| + C \leq C 2^{j/2}. \end{aligned} \quad (6.7)$$

Using (6.7) and Proposition 6.1, we obtain

$$\mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^4 \right) \leq C 2^j \mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^2 \right) \leq C 2^j \frac{1}{n}.$$

The proof of Proposition 6.2 is complete. □

**Proposition 6.3** *Suppose that the assumptions of Section 2 hold. For any  $j \in \{j_0, \dots, j_1\}$  and any  $k \in \{0, \dots, 2^j - 1\}$ , let  $\beta_{j,k} = \int_0^1 F(G^{-1}(x)) \psi_{j,k}(x) dx$ ,  $\hat{\beta}_{j,k}$  be (4.1) and  $\rho_n$  be defined as in (4.4). Then there exist two constants,  $\kappa > 0$  and  $C > 0$ , such that*

$$\mathbb{P} \left( |\hat{\beta}_{j,k} - \beta_{j,k}| \geq \kappa \rho_n / 2 \right) \leq C \frac{1}{n^4}.$$

**Proof of Proposition 6.3.** We shall use the Bernstein inequality for exponentially strongly mixing process presented in Lemma 6.1 below. The proof can be found in [21].



**Lemma 6.1** ([21]) *Let  $\gamma > 0$ ,  $c > 0$  and  $(Z_i)_{i \in \mathbb{Z}}$  be a strictly stationary process defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  with the  $m$ -th strongly mixing coefficient (2.1). Let  $n$  be a positive integer,  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function and, for any  $i \in \mathbb{Z}$ ,  $V_i = h(Z_i)$ . We assume that  $E(V_1) = 0$  and there exists a constant  $M > 0$  satisfying  $|V_1| \leq M$ . Then, for any  $m \in \{1, \dots, n\}$  and any  $\lambda > 0$ , we have*

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n V_i \right| \geq \lambda \right) \leq 4 \exp \left( -\frac{\lambda^2 n}{16(D_m/m + \lambda M m/3)} \right) + 32 \frac{M}{\lambda} n a_m,$$

where  $D_m = \max_{l \in \{1, \dots, 2m\}} \text{Var} \left( \sum_{i=1}^l V_i \right)$ .

For any  $i \in \{1, \dots, n\}$ , set

$$V_i = \delta_i \psi_{j,k}(G(U_i)) - \beta_{j,k}.$$

Then

$$\mathbb{P} \left( |\hat{\beta}_{j,k} - \beta_{j,k}| \geq \kappa \rho_n / 2 \right) = \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n V_i \right| \geq \kappa \rho_n / 2 \right), \quad (6.8)$$

$V_1, \dots, V_n$  are identically distributed, depend on the strictly stationary strongly mixing process  $(\delta_i, U_i)_{i \in \mathbb{Z}}$  satisfying (2.2), Propositions 6.1 and 6.2 give

$$E(V_1) = 0, \quad E(V_1^2) \leq E \left( (\delta_1 \phi_{j,k}(G(U_1)))^2 \right) \leq 1,$$

and, using similar arguments to (6.7),  $|V_1| \leq C 2^{j/2} \leq C 2^{j_1/2} \leq C(n/(\ln n)^3)^{1/2}$ . Now set  $m = \lfloor u \ln n \rfloor$  with  $u > 0$  (chosen later). Proceeding as in the bounds of **S** and **T**<sub>1</sub> in the proof of Proposition 6.1 with  $l$  instead of  $n$ , we have

$$\begin{aligned} \max_{l \in \{1, \dots, 2m\}} \text{Var} \left( \sum_{i=1}^l V_i \right) &\leq C \max_{l \in \{1, \dots, 2m\}} (l + l^2 2^{-j}) \leq C \max_{l \in \{1, \dots, 2m\}} (l + l^2 2^{-j_0}) \\ &\leq C \left( m + \frac{m^2}{\ln n} \right) \leq C m. \end{aligned}$$

It follows from Lemma 6.1 applied with these  $V_1, \dots, V_n$ ,  $\lambda = \kappa C \rho_n$ ,  $\rho_n = (\ln n/n)^{1/2}$ ,  $m = u \ln n$  with  $u > 0$  (chosen later),  $M = C(n/(\ln n)^3)^{1/2}$  and (2.2) that

$$\begin{aligned} &\mathbb{P} \left( |\hat{\beta}_{j,k} - \beta_{j,k}| \geq \kappa \rho_n / 2 \right) \\ &\leq C \left( \exp \left( -C \frac{\kappa^2 \rho_n^2 n}{1 + \kappa \rho_n m M} \right) + \frac{n^{1/2}}{\rho_n (\ln n)^{3/2}} n \exp(-cm) \right) \\ &\leq C \left( \exp \left( -C \frac{\kappa^2 \ln n}{1 + \kappa (\ln n/n)^{1/2} u \ln n (n/(\ln n)^3)^{1/2}} \right) + n^2 \exp(-cu \ln n) \right) \\ &\leq C \left( n^{-C\kappa^2/(1+\kappa u)} + n^{2-cu} \right). \end{aligned}$$

Taking  $\kappa$  and  $u$  large enough, we obtain

$$\mathbb{P} \left( |\hat{\beta}_{j,k} - \beta_{j,k}| \geq \kappa \rho_n / 2 \right) \leq C \frac{1}{n^4}.$$

This ends the proof of Proposition 6.3.

□

## 6.2 Proof of Theorem 5.1

**Proof of Theorem 5.1.** Since  $F \in \mathbb{L}_2([0, 1])$ , we can write

$$F(x) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k} \phi_{j_0,k}(G(x)) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(G(x)), \quad x \in [0, 1],$$

where  $\alpha_{j_0,k} = \int_0^1 F(G^{-1}(x)) \phi_{j_0,k}(x) dx$  and  $\beta_{j,k} = \int_0^1 F(G^{-1}(x)) \psi_{j,k}(x) dx$ .

We have, for any  $x \in [0, 1]$ ,

$$\begin{aligned} & \hat{F}(x) - F(x) \\ &= \sum_{k=0}^{2^{j_0}-1} (\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \phi_{j_0,k}(G(x)) + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left( \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa \rho_n\}} - \beta_{j,k} \right) \psi_{j,k}(G(x)) \\ & - \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(G(x)). \end{aligned}$$

We now need the following lemma which is an immediate consequence of [20, Lemma 2] and the Minkowski inequality.

**Lemma 6.2 ([20])** *Suppose that (2.3) holds. Then, for any sequences  $(u_{j,k}) \in \ell_2(\mathbb{N}^2)$  and any integers  $j_0$  and  $j_1$  such that  $j_1 > j_0 \geq j_0$ , there exists a constant  $C > 0$  such that*

$$\int_0^1 \left( \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} u_{j,k} \psi_{j,k}(G(x)) \right)^2 dx \leq C \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} u_{j,k}^2 w_{j,k} \right)^{1/2} \right)^2,$$

where  $w_{j,k}$  is defined by (3.2).

It follows from Lemma 6.2 and the Cauchy-Schwarz inequality that

$$\mathbb{E} \left( \int_0^1 \left( \hat{F}(x) - F(x) \right)^2 dx \right) \leq C(\mathbf{F} + \mathbf{G} + \mathbf{H}), \quad (6.9)$$

where

$$\begin{aligned} \mathbf{F} &= 2^{j_0} \sum_{k=0}^{2^{j_0}-1} \mathbb{E} \left( (\hat{\alpha}_{j_0,k} - \alpha_{j_0,k})^2 \right) w_{j_0,k}, \\ \mathbf{G} &= \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbb{E} \left( \left( \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa \rho_n\}} - \beta_{j,k} \right)^2 \right) w_{j,k} \right)^{1/2} \right)^2 \end{aligned}$$

and

$$\mathbf{H} = \left( \sum_{j=j_1+1}^{\infty} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \beta_{j,k}^2 w_{j,k} \right)^{1/2} \right)^2.$$

Using Proposition 6.1 and  $\sum_{k=0}^{2^{j_0}-1} w_{j_0,k} = 1$ , we obtain

$$\mathbf{F} \leq C \frac{1}{n} 2^{j_0} \sum_{k=0}^{2^{j_0}-1} w_{j_0,k} \leq C \frac{\ln n}{n} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \quad (6.10)$$

Since  $F \in B_{s,\infty}^w(M)$ , we have

$$\begin{aligned} \mathbf{H} &\leq C \left( \sum_{j=j_1+1}^{\infty} 2^{-js} \right)^2 \leq C 2^{-2j_1 s} \leq C \left( \frac{(\ln n)^3}{n} \right)^{2s} \\ &\leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \end{aligned} \quad (6.11)$$

Let us now bound the term  $\mathbf{G}$ . Observe that

$$\mathbf{G} \leq C(\mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3 + \mathbf{G}_4), \quad (6.12)$$

where

$$\begin{aligned} \mathbf{G}_1 &= \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa \rho_n\}} \mathbf{1}_{\{|\beta_{j,k}| < \kappa \rho_n/2\}} \right) w_{j,k} \right)^{1/2} \right)^2, \\ \mathbf{G}_2 &= \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa \rho_n\}} \mathbf{1}_{\{|\beta_{j,k}| \geq \kappa \rho_n/2\}} \right) w_{j,k} \right)^{1/2} \right)^2, \\ \mathbf{G}_3 &= \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbb{E} \left( \beta_{j,k}^2 \mathbf{1}_{\{|\hat{\beta}_{j,k}| < \kappa \rho_n\}} \mathbf{1}_{\{|\beta_{j,k}| \geq 2\kappa \rho_n\}} \right) w_{j,k} \right)^{1/2} \right)^2 \end{aligned}$$

and

$$\mathbf{G}_4 = \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbb{E} \left( \beta_{j,k}^2 \mathbf{1}_{\{|\hat{\beta}_{j,k}| < \kappa \rho_n\}} \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa \rho_n\}} \right) w_{j,k} \right)^{1/2} \right)^2.$$

*Upper bounds for  $\mathbf{G}_1 + \mathbf{G}_3$ .* Note that

$$\begin{aligned} \left\{ |\hat{\beta}_{j,k}| < \kappa \rho_n, |\beta_{j,k}| \geq 2\kappa \rho_n \right\} &\subseteq \left\{ |\hat{\beta}_{j,k} - \beta_{j,k}| > \kappa \rho_n/2 \right\}, \\ \left\{ |\hat{\beta}_{j,k}| \geq \kappa \rho_n, |\beta_{j,k}| < \kappa \rho_n/2 \right\} &\subseteq \left\{ |\hat{\beta}_{j,k} - \beta_{j,k}| > \kappa \rho_n/2 \right\} \end{aligned}$$

and

$$\left\{ |\hat{\beta}_{j,k}| < \kappa \rho_n, |\beta_{j,k}| \geq 2\kappa \rho_n \right\} \subseteq \left\{ |\beta_{j,k}| \leq 2|\hat{\beta}_{j,k} - \beta_{j,k}| \right\}.$$

So

$$\mathbf{G}_1 + \mathbf{G}_3 \leq C \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| > \kappa \rho_n/2\}} \right) w_{j,k} \right)^{1/2} \right)^2.$$

It follows from the Cauchy-Schwarz inequality, Proposition 6.2, Proposition 6.3 and  $2^j \leq 2^{j_1} \leq n$  that

$$\begin{aligned} &\mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^2 \mathbf{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| > \kappa \rho_n/2\}} \right) \\ &\leq \left( \mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^4 \right) \right)^{1/2} \left( \mathbb{P} \left( |\hat{\beta}_{j,k} - \beta_{j,k}| > \kappa \rho_n/2 \right) \right)^{1/2} \\ &\leq C \left( 2^j \frac{1}{n} \right)^{1/2} \left( \frac{1}{n^4} \right)^{1/2} \leq C \frac{1}{n^2}. \end{aligned}$$

Since  $\sum_{k=0}^{2^j-1} w_{j,k} = 1$ , we have

$$\begin{aligned} \mathbf{G}_1 + \mathbf{G}_3 &\leq C \frac{1}{n^2} \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} w_{j,k} \right)^{1/2} \right)^2 = C \frac{1}{n^2} \left( \sum_{j=j_0}^{j_1} 2^{j/2} \right)^2 \\ &\leq C \frac{1}{n^2} 2^{j_1} \leq C \frac{1}{n} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \end{aligned} \quad (6.13)$$

Upper bound for  $\mathbf{G}_2$ . Using again Proposition 6.2, we obtain

$$\mathbb{E} \left( \left( \hat{\beta}_{j,k} - \beta_{j,k} \right)^2 \right) \leq C \frac{1}{n} \leq C \frac{\ln n}{n}.$$

Hence

$$\mathbf{G}_2 \leq C \frac{\ln n}{n} \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}| > \kappa \rho_n / 2\}} w_{j,k} \right)^{1/2} \right)^2.$$

Let  $j_2$  be the integer defined by

$$\frac{1}{2} \left( \frac{n}{\ln n} \right)^{1/(2s+1)} < 2^{j_2} \leq \left( \frac{n}{\ln n} \right)^{1/(2s+1)}. \quad (6.14)$$

We have

$$\mathbf{G}_2 \leq C(\mathbf{G}_{2,1} + \mathbf{G}_{2,2}),$$

where

$$\mathbf{G}_{2,1} = \frac{\ln n}{n} \left( \sum_{j=j_0}^{j_2} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}| > \kappa \rho_n / 2\}} w_{j,k} \right)^{1/2} \right)^2$$

and

$$\mathbf{G}_{2,2} = \frac{\ln n}{n} \left( \sum_{j=j_2+1}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}| > \kappa \rho_n / 2\}} w_{j,k} \right)^{1/2} \right)^2.$$

Using  $\mathbf{1}_{\{|\beta_{j,k}| > \kappa \rho_n / 2\}} \leq 1$  and  $\sum_{k=0}^{2^j-1} w_{j,k} = 1$ ,

$$\begin{aligned} \mathbf{G}_{2,1} &\leq C \frac{\ln n}{n} \left( \sum_{j=j_0}^{j_2} 2^{j/2} \left( \sum_{k=0}^{2^j-1} w_{j,k} \right)^{1/2} \right)^2 = C \frac{\ln n}{n} \left( \sum_{j=j_0}^{j_2} 2^{j/2} \right)^2 \\ &\leq C \frac{\ln n}{n} 2^{j_2} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)} \end{aligned}$$

and, since  $F \in B_{s,\infty}^w(M)$ ,

$$\begin{aligned} \mathbf{G}_{2,2} &\leq C \frac{\ln n}{n \rho_n^2} \left( \sum_{j=j_2+1}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \beta_{j,k}^2 w_{j,k} \right)^{1/2} \right)^2 \leq C \left( \sum_{j=j_2+1}^{j_1} 2^{-js} \right)^2 \\ &\leq C 2^{-2j_2 s} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

So

$$\mathbf{G}_2 \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \quad (6.15)$$

Upper bound for  $\mathbf{G}_4$ . We have

$$\mathbf{G}_4 \leq \left( \sum_{j=j_0}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa\rho_n\}} w_{j,k} \right)^{1/2} \right)^2.$$

Let  $j_2$  be the integer (6.14). Then

$$\mathbf{G}_4 \leq C(\mathbf{G}_{4,1} + \mathbf{G}_{4,2}),$$

where

$$\mathbf{G}_{4,1} = \left( \sum_{j=j_0}^{j_2} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa\rho_n\}} w_{j,k} \right)^{1/2} \right)^2$$

and

$$\mathbf{G}_{4,2} = \left( \sum_{j=j_2+1}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa\rho_n\}} w_{j,k} \right)^{1/2} \right)^2.$$

Using  $\beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa\rho_n\}} \leq C\rho_n^2$  and  $\sum_{k=0}^{2^j-1} w_{j,k} = 1$ , we have

$$\begin{aligned} \mathbf{G}_{4,1} &\leq C\rho_n^2 \left( \sum_{j=j_0}^{j_2} 2^{j/2} \left( \sum_{k=0}^{2^j-1} w_{j,k} \right)^{1/2} \right)^2 = C \frac{\ln n}{n} \left( \sum_{j=j_0}^{j_2} 2^{j/2} \right)^2 \\ &\leq C \frac{\ln n}{n} 2^{j_2} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

Since  $F \in B_{s,\infty}^w(M)$ , we have

$$\begin{aligned} \mathbf{G}_{4,2} &\leq \left( \sum_{j=j_2+1}^{j_1} 2^{j/2} \left( \sum_{k=0}^{2^j-1} \beta_{j,k}^2 w_{j,k} \right)^{1/2} \right)^2 \leq C \left( \sum_{j=j_2+1}^{j_1} 2^{-js} \right)^2 \\ &\leq C 2^{-2j_2 s} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

So

$$\mathbf{G}_4 \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \quad (6.16)$$

It follows from (6.12), (6.13), (6.15) and (6.16) that

$$\mathbf{G} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}. \quad (6.17)$$

Combining (6.9), (6.10), (6.11) and (6.17), we have

$$\mathbb{E} \left( \int_0^1 \left( \hat{F}(x) - F(x) \right)^2 dx \right) \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 5.1 is complete.

□

**Acknowledgments.** The authors would like to thank Yves Rozenholc for providing us programs for the implementation of the piecewise polynomial regression estimator, and the reviewer for helping us to improve the paper.

## References

- [1] Autin, F. (2008). Maxisets for mu-thresholding rules (2008). *Test*, 17, 2, 332-349.
- [2] Brunel, E. and Comte, F. (2009). Cumulative distribution function estimation under interval censoring case 1. *Electron. J. Stat.*, 3, 1-24.
- [3] Brutti, P. (2008). Warped wavelets and vertical thresholding. Preprint. arXiv:0801.3319v1.
- [4] Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.
- [5] Chagny, G. (2011). Penalization versus Goldenshluger-Lepski strategies in warped bases regression, *ESAIM P&S*, to appear.
- [6] Chagny, G. (2012a). Warped bases for conditional density estimation. preprint MAP5, hal-00641560.
- [7] Chagny, G. (2012b). Nonparametric warped kernel estimators. preprint. MAP5, hal-00715184.
- [8] Chesneau, C. (2007). A maxiset approach of a Gaussian noise model. (2007). *TEST*, 16, 3, 523-546.
- [9] Chesneau, C. and Willer, T. (2007). Numerical performances of a warped wavelet estimation procedure for regression in random design, Preprint, HAL:hal-00133831.
- [10] Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24, 1, 54-81.
- [11] Daubechies, I., (1992) Ten lectures on wavelets, SIAM, 1992.
- [12] Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.
- [13] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Applied Computational Harmonic Analysis*, 3, 215-228.
- [14] Diamond, I.D. and McDonald, J.W. (1991). The analysis of current status data. *Demographic applications of event history analysis*. (eds J. Trussell, R. Hankinson & J. Tilton). Oxford University Press, Oxford.
- [15] Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, 24, 508-539.
- [16] Doukhan, P. (1994). *Mixing. Properties and Examples*. Lecture Notes in Statistics 85. Springer Verlag, New York.
- [Dunson and Dinse] Dunson, D. B. and Dinse, G. E. (2002). Bayesian models for multivariate current status data with informative censoring. *Biometrics*, 58, 79-88.

- [18] Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*. Lectures Notes in Statistics New York 129, Springer Verlag.
- [19] Jewell, N. P. and van der Laan, M. (2004). *Current status data: review, recent developments and open problems*. Advances in survival analysis, 625-642, Handbook of Statist., 23, Elsevier, Amsterdam.
- [20] Kerkycharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli*, 10(6) :1053-1105.
- [21] Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statist. Decisions*, 19, (1), 9-26.
- [22] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.
- [23] Modha, D. and Masry, E. (1996). Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inform. Theory*, 42, 2133-2145.
- [24] Pham Ngoc, T. M. (2009). Regression in random design and Bayesian warped wavelets estimators. *Electronic Journal of Statistics*, **3**, 1084-1112.
- [25] van der Vaart, A. and van der Laan, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.*, 2, Art 9, 42pp.
- [26] Withers, C. S. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 477-480.